INTENTIONAL-STANCE CHARACTERIZATION OF A GENERAL CONTENT STREAM OR REPOSITORY

RELATED APPLICATION DATA

This application is related to U.S. Patent No. 6,108,619, issued August 22, 2000, titled "METHOD AND APPARATUS FOR SEMANTIC CHARACTERIZATION," co-pending U.S. Patent application Serial No. 09/512,963, titled "CONSTRUCTION, MANIPULATION, AND COMPARISON OF A MULTI-DIMENSIONAL SEMANTIC SPACE," filed February 25, 2000, and to co-pending U.S. Patent application Serial No. _____, titled "A METHOD AND MECHANISM FOR THE CREATION, MAINTENANCE, AND COMPARISON OF SEMANTIC ABSTRACTS," filed _____, all commonly assigned.

FIELD OF THE INVENTION

This invention pertains to determining in a computer the semantic content of documents, and more particularly to determining and responding to the emotional content of the documents.

BACKGROUND OF THE INVENTION

U.S. Patent applications Serial No. 09/109,804, titled "METHOD AND APPARATUS FOR SEMANTIC CHARACTERIZATION," filed July 2, 1998, U.S. Patent application Serial No. 09/512,963, titled "CONSTRUCTION, MANIPULATION, AND COMPARISON OF A MULTI-DIMENSIONAL SEMANTIC SPACE," filed February 25, 2000, and U.S. Patent application Serial No. _____, titled "A METHOD AND MECHANISM FOR THE CREATION, MAINTENANCE, AND COMPARISON OF SEMANTIC ABSTRACTS," filed _____ provide a framework and mechanism in a computer for providing a "quantized" meaning-representation. But the semantic meaning of documents is not the only factor controlling interpretation of the documents. People have emotional responses to the world around them. What one person might view with emotional detachment can cause another person to respond strongly. Even more, people can have different emotional responses to the same content, depending on their point of view. For

example, given an article proposing abortion regulation, many people would vehemently oppose the regulation, while others would avidly support it. Currently, however, no mechanism exists to enable users to automate the process of identifying and responding to documents or other content to which the users have an emotional response.

The present invention addresses these and other problems associated with the prior art.

SUMMARY OF THE INVENTION

The invention is a computer-implemented method and apparatus for building a template specifying an emotional response to a content stream. A dictionary is selected. The dictionary includes concepts organized as a directed set and chains overlaid on top of the concepts. A subset of the chains are selected as intentional stance basis chains (ISBCs). State vectors are constructed in a topological vector space for a subset of the concepts in the dictionary. The state vectors are assembled into a template, and an action and threshold distance are assigned to, or associated with, the template. As content becomes available, impact summaries of the content are compared with the template. If the impact summary is within the threshold distance of the template, the action is performed.

The invention enables a computer to determine whether a content stream provokes an emotional response. A template is constructed in a topological vector space as described above. An impact summary is constructed or generated for the content stream, and is compared with the template. The comparison can be made by measuring a distance between the impact summary and the template in the topological vector space.

The foregoing and other features, objects, and advantages of the invention will become more readily apparent from the following detailed description, which proceeds with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a computer system on which the invention can operate to create templates and impact summaries for content streams.

FIG. 2 shows a dictionary including concepts over which chains have been laid, from which intentional stance basis chains can be selected.

FIGs. 3A-3G show eight different intentional stance basis chains in the dictionary of FIG. 2.

FIG. 4 shows a two-dimensional topological vector space in which a template is presented.

FIG. 5 shows a two-dimensional topological vector space in which the emotional impact of a document is determined by measuring the distance between an impact summary for the document and a template.

FIG. 6 shows a flowchart of a method to construct it a template in the computer system of FIG. 1 according to the preferred embodiment of the invention.

FIG. 7 shows a flowchart of a method to analyze a content stream in the computer system of FIG. 1 to determine if an emotional response is required.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Introduction to Templates

U.S. Patent application Serial No. 09/512,963, titled "CONSTRUCTION, MANIPULATION, AND COMPARISON OF A MULTI-DIMENSIONAL SEMANTIC SPACE," filed February 25, 2000, incorporated by reference herein and referred to as "the Construction application," describes the creation of state vectors in a topological vector space for concepts in a dictionary. U.S. Patent application Serial No. _____, titled "A METHOD AND MECHANISM FOR THE CREATION, MAINTENANCE, AND COMPARISON OF SEMANTIC ABSTRACTS," filed _____, incorporated by reference herein and referred to as "the Semantic Abstracts application," takes state vectors for individual lexemes/lexeme phrases and assembles them into semantic abstracts for documents.

The techniques used in the Semantic Abstracts application can be applied to building intentional stance impact templates (or "templates" for short). Instead of working with an abstract of a document, however, the template begins with an abstract of content to which the user would have an emotional response. For example, an anti-abortion activist might select the phrase "The doctor aborted the fetus" as an abstract of content to which he has an emotional response. The abstract is then parsed into its constituent lexemes/lexeme phrases, and each lexeme/lexeme phrase is mapped to a state vector in a topological vector space. The state vectors are collected to form the template, and an action is assigned to the template. For example, the anti-abortion activist might assign as an action to the template to send a letter to each of his representatives in Congress. (Although example uses of the invention will be presented using textual documents, a person skilled in the art will recognize that the invention

is equally applicable to non-textual works. For example, images or music can provoke emotional responses in people just as strongly as written words.)

Alternatively, the hypothetical anti-abortion activist could be reading an electronic version of a report. Upon encountering a paragraph that the activist finds incendiary, the activist can mark the paragraph. The paragraph can be semantically analyzed and disassembled into its component lexemes/lexeme phrases. Vectors can be created from the lexemes/lexeme phrases. If necessary, the vectors can be filtered. Then the filtered vectors can be assembled into a template, and an action assigned to the template. In this manner, the template can be created without the activist having to summarize the material.

A reader familiar with the Semantic Abstracts application will observe that templates have similarities with semantic abstracts. Both represent distilled content. Both include sets of state vectors in topological vector spaces. But whereas semantic abstracts are associated with particular documents, templates are not linked to any particular document.

Systems for Building and Using Templates

Now that the reader is familiar with the high-level design goals of templates, the techniques of constructing and using templates can be explained in greater detail. FIG. 1 shows a computer system 105 on which a method and apparatus for using a multi-dimensional semantic space can operate. Computer system 105 conventionally includes a computer 110, a monitor 115, a keyboard 120, and a mouse 125. Computer system 105 can also be an Internet appliance, lacking monitor 115, keyboard 120, or mouse 125. Optional equipment not shown in FIG. 1 can include a printer and other input/output devices. Also not shown in FIG. 1 are the conventional internal components of computer system 105: e.g., a central processing unit, memory, file system, etc.

Computer system 105 further includes software 130. In FIG. 1, software 130 includes dictionary 135, intentional stance basis chains (ISBCs) 140, template 145, action 150, and comparison software 155. Dictionary 135 provides the foundation for the topological vector space used to construct the template. FIG. 2 shows a sample dictionary 135 including concepts and chains. (Dictionary 135 shown in FIG. 2 is drawn from the Construction application.) Dictionary 135 includes a set of concepts, typically organized as a directed set. At the top of the directed set is a particular concept known as the maximal element. For each concept in the directed set other than the maximal element, there is at least one "parent" concept in the directed set that is a generalization of that concept. (There can be multiple

"parent" concepts, because language allows for overloading of words.) The relationships between concepts can be extended all the way to the maximal element; the hierarchy of such relationships between the maximal element and each concept are called chains.

ISBCs 140 are a selected subset of the chains in dictionary 135. For example, FIGs. 3A-7G show eight different chains in the dictionary of FIG. 2 that can be selected as ISBCs 140. FIG. 3A shows chain 305, which extends to concept "man" through concept "energy." FIG. 3B shows chain 310 extending to concept "iguana." FIG. 3C shows another chain 315 extending to concept "man" via a different path. FIGs. 3D-3G show other chains.

Returning to FIG. 1, ISBCs 140 play the same role in building state vectors for templates that the basis chains play in creating state vectors in the Construction application. Any desired set of ISBCs can be selected from the chains in dictionary 135. Although comparing a template with an impact summary requires that the template and impact summary be constructed using the same selected chains, topological vector space transformations can be performed to allow comparison. One of the impact summary and the template can be transformed to use the other's basis chains, or both can be transformed to a normative, preferred dictionary/basis combination.

Once ISBCs 140 have been selected, the lexemes/lexeme phrases chosen to create the template are mapped to state vectors in a topological vector space. These state vectors are then assembled into template 145. Action 150 representing the user's desired response is assigned to template 145. Finally, comparison software 155 is used to compare template 145 with an impact summary for the content in question. If necessary, comparison software 155 can construct an impact summary for the content.

Although the content compared with template 145 can be found stored on computer system 105, this is not required. FIG. 1 shows computer system 105 accessing a content stream 160 over network connection 165. Content stream 160 can be a single document, or it can include multiple sources. Content streams with multiple sources are common in today's world. For example, newsgroups and discussion lists (e-mail lists) allow multiple users to carry on several conversations on multiple topics at the same time. Newsgroups and discussion lists are typically organized into a hierarchy. The newsgroup itself has a content focus. This content focus is divided into sub-topics, called threads. Each thread is further divided into individual messages from individual users. FIG. 1 shows content stream 160 as having two threads. Thread one has two messages, and thread two has three messages.

Depending on the level of abstraction the user wishes to track, the template can be compared with different impact summaries. For example, a user may have an emotional reaction to the entire content stream, without looking at individual messages within the content stream (e.g., an entire newsgroup devoted to support of abortion). Alternatively, the user might have a reaction to only a thread within the newsgroup (e.g., a thread about abortion within a medical newsgroup), or to only an individual message.

Network connection 165 can include any kind of network connection. For example, network connection 165 can enable computer system 105 to access content stream 160 over a local area network (LAN), a wide area network (WAN), a global internetwork, a wireless network or broadcast network, or any other type of network. Similarly, once collected, the impact summary can be stored somewhere on computer system 105, or can be stored elsewhere using network connection 165.

Using Templates

FIG. 4 shows a two-dimensional topological vector space in which a template includes a set of state vectors. (FIGs. 4 and 5, although accurate representations of a topological vector space, are greatly simplified for example purposes, since most topological vector spaces will have significantly higher dimensions.) In FIG. 4, template 405 includes the state vectors represented by the "x" symbols. (For clarity, the line segments from the origin of the topological vector space to the heads of the state vectors are not shown in FIG. 4.)

Circle 410 represents the threshold distance defined for template 405 before an action is taken. The reader will recognize that circle 410 is an abstraction, since in the preferred embodiment distance is not measured from a single point in the topological vector space. Instead, in the preferred embodiment distance is measured from the entire set of vectors comprising the template. But if template 405 could be reduced to a single point in the topological vector space, circle 410 could represent a threshold distance. Any impact summaries that came within circle 410 would then trigger the action associated with template 405.

Although in the preferred embodiment the template includes only one associated action and one threshold distance, a user may have different emotional responses depending on how close the content source gets to the template. For example, an anti-abortion activist might find a theoretical discussion about the benefits of abortion merely unsettling and worth

only paying closer attention (say, having follow-up articles e-mailed to his attention). An article reporting an abortion performed by a doctor might make the activist feel quite unhappy (prompting a letter to his congressperson). And a report about pending legislation to keep abortion legal might make the activist very angry (and cause him to start circulating a petition to prevent the legislation). The activist might then set up a number of different actions depending on how close the content source gets to the template, and associate different threshold distances to each action. Circle 415 represents such a second threshold distance associated with a different action for the template.

The Hausdorff Distance Function

As template 405 consists of a set of vectors, the preferred embodiment for measuring distance from template 405 to an impact summary is via the Hausdorff distance function. The following is excerpted from the Construction and Semantic Abstracts applications.

Recall that in the Construction application it was shown that $\mathcal{H}(\mathbf{S})$ is the set of all compact (non-empty) subsets of a metrizable space $\mathbf{S}$. The Hausdorff distance h is defined as follows: Define the pseudo-distance $\xi(x, u)$ between the point $x \in \mathbf{S}$ and the set $u \in \mathcal{H}(\mathbf{S})$ as

$$\xi(x, u) = \min\{d(x, y) : y \in u\}.$$

Using ξ define another pseudo-distance $\lambda(u, v)$ from the set $u \in \mathcal{H}(\mathbf{S})$ to the set $v \in \mathcal{H}(\mathbf{S})$:

$$\lambda(u, v) = \max\{\xi(x, v) : x \in u\}.$$

Note that in general it is *not* true that $\lambda(u, v) = \lambda(v, u)$. Finally, define the distance $h(u, v)$ between the two sets $u, v \in \mathcal{H}(\mathbf{S})$ as

$$h(u, v) = \max\{\lambda(u, v), \lambda(v, u)\}.$$

The distance function h is called the *Hausdorff* distance. Note that

$h(u, v) = h(v, u)$,

$0 < h(u, v) < \infty$ for all $u, v \in \mathcal{H}(\mathbf{S})$, $u \neq v$,

$h(u, u) = 0$ for all $u \in \mathcal{H}(\mathbf{S})$, and

$h(u, v) \leq h(u, w) + h(w, v)$ for all $u, v, w \in \mathcal{H}(\mathbf{S})$.

Alternatives to the Hausdorff Distance Function

In the Semantic Abstracts application, one alternative to using the Hausdorff distance function was to locate a centroid vector for the semantic abstract. The same techniques are applicable to templates. By locating a centroid vector (using any measure of central tendency) for the template, distance functions that depend on single points (e.g., Euclidean distance) can be used to measure distance from the template to the impact summary.

Measuring Distance from the Template

FIG. 5 shows a two-dimensional topological vector space in which template 405 is compared with an impact summary for a document. (To avoid clutter in the drawing, FIG. 5 shows template 405 and impact summary 505 in different graphs of the same topological vector space. The reader can imagine the template and impact summary as being in the same graph.) Using the Hausdorff distance function h, the distance 510 between template 405 and impact summary 505 can be quantified. If distance 510 is smaller than the threshold distance defined for template 405, then the action associated with template 405 will be triggered.

Procedural Implementation

FIG. 6 shows a flowchart of a method to construct a template in the computer system of FIG. 1 according to the preferred embodiment of the invention. At step 605, a dictionary (a directed set of concepts and chains) is selected. At step 610, a subset of the chains in the dictionary is selected as ISBCs. Any dictionary can and basis chains can be used, but one dictionary/basis combination might be preferred over another, for example, because it would avoid requiring a topological vector space transformation. At step 615, the concepts (lexemes/lexeme phrases) in the dictionary that generate an emotional response are selected. The concepts are typically selected based on trigger concepts, often in phrases that make sense to the reader (for example, "doctors performing abortions at clinics"). At step 620, state vectors are constructed for the selected concepts. The ISBCs are used to construct the state vectors. At step 625, the state vectors are assembled (collected into a set) to form the template. Finally, at step 630, an action and threshold distance are assigned to the template.

FIG. 7 shows a flowchart of a method to analyze a content stream in the computer system of FIG. 1 to determine if an emotional response is triggered. At step 705, an impact summary is constructed for the content stream. At step 710, the distance from the impact summary to the template is measured. At step 715, the distance is checked against the

threshold for taking the action associated with the template. If the distance indicates that the impact summary is within the threshold for taking action associated with the template, then at step 720 the action is performed.

Monitoring Changes in Content

Although the above description treats content as a static object, it is not. Content changes over time. For example, returning to the example of the newsgroup, threads die out as users stop posting new messages regarding the thread or moderators kill improper threads. New threads pop up as new subjects are proposed. People's viewpoints change as one argument or another sways them. As content changes, the need for action can accordingly change. For example, if the newsgroup is about medicine, the above-described anti-abortion activist might originally have no emotional response one way or the other to the newsgroup. But when a thread on the subject of abortion appears, the activist's interest increases. And as people start posting messages supporting abortion, the activist begins to take action. Accordingly, the user needs to update impact summaries to make sure the impact summary is current. In FIG. 7, dashed line 725 reflects this need. Periodically, the system starts over at step 705 to determine the current content of the source.

A Benefit of Templates

One use of the intentional stance impact template is for authors to try and gauge how their works will be received. An author can construct a template he expects his readers to use in reviewing the work. The author can then apply the template to the work and see how close the work comes to the template. For example, if the work discusses the medical benefits of abortion, the author can construct a template that an anti-abortion activist might use, and see how close the article comes to the template. This allows the author to estimate how others might respond to the work. The author can then further refine the work if the provoked response is not as desired.

Another example of how templates can be used is illustrated in related co-pending U.S. Patent Application Serial No. _____, entitled "POLICY ENFORCEMENT USING THE SEMANTIC CHARACTERIZATION OF TRAFFIC," filed simultaneously herewith.

Having illustrated and described the principles of our invention in a preferred embodiment thereof, it should be readily apparent to those skilled in the art that the invention

can be modified in arrangement and detail without departing from such principles. We claim all modifications coming within the spirit and scope of the accompanying claims.